

Error rates of Belavkin weighted quantum measurements and a converse to Holevo's asymptotic optimality Theorem

Jon Tyson*

Jefferson Laboratory, Harvard University, Cambridge MA 02138, USA

November 12, 2008

Journal Ref: Phys. Rev. A **79**, 032343 (2009).

Abstract

We compare several instances of pure-state Belavkin weighted square-root measurements from the standpoint of minimum-error discrimination of quantum states. The quadratically weighted measurement is proven superior to the so-called “pretty good measurement” (PGM) in a number of respects:

1. Holevo's quadratic weighting unconditionally outperforms the PGM in the case of two-state ensembles, with equality only in trivial cases.
2. A converse of a theorem of Holevo is proven, showing that a weighted measurement is asymptotically-optimal only if it is quadratically weighted.

Counter-examples for three states are constructed. The cube-weighted measurement of Ballester, Wehner, and Winter is also considered. Sufficient optimality conditions for various weights are compared.

*jonetyson@X.Y.Z, where X=post, Y=Harvard, Z=edu

1 Introduction

1.1 Optimal measurements

Consider an ensemble $\mathcal{E}_m^{\text{mixed}}$ of mixed quantum states ρ_k with *a priori* probabilities p_k , $k = 1, \dots, m$, and unit normalizations $\text{Tr } \rho_k = 1$. Of fundamental importance is

The minimum-error quantum distinguishability problem: If an unknown state ρ_k is blindly drawn from the ensemble, what is the chance that the corresponding value of k may be correctly identified by performing an optimally chosen quantum measurement?

The modern approach to this problem is to consider measurements defined by

Definition 1 A *positive-operator valued measure (POVM)* $\{M_k\}$ (see, for example, p. 74 of [1]) is a collection of positive semidefinite operators on a Hilbert space \mathcal{H} such that $\sum M_k = \mathbb{1}$. The probability that the value i is detected when the POVM is applied to the state ρ_j is given by $p_{i|j} = \text{Tr } M_i \rho_j$. In particular, the **success rate** for the POVM to distinguish the ensemble $\mathcal{E}_m^{\text{mixed}}$ is given by

$$P_{\text{succ}} = \sum_{k=1}^m p_k \text{Tr}(\rho_k M_k). \quad (1)$$

Minimum-error quantum measurement was first considered in the 1960s in the design of high performance optical detectors [1]. More recently, this problem has been fundamentally important in quantum Shannon theory (for example [2–4]) and in construction of quantum algorithms for the *Hidden Subgroup Problem*. [5–10]. Various necessary and sufficient conditions for optimal measurements have been derived [11–16] (see also [17]). A number of relatively recent works give interesting general upper and/or lower bounds on the quantum distinguishability problem. [2, 18–25] Explicitly solving the general optimal measurement problem is most likely impossible, but in specific numerical cases one may compute the optimal measurement by numerical iteration [25–28] or by numerical solution of the associated semidefinite program [17].

The optimal measurement problem has been generalized to wave discrimination [29] and to optimal reversals of quantum channels, in the sense of average entanglement fidelity [19, 30–34]. More recently, the success-rate of optimal measurements has been expressed in terms of the conditional min-entropy of corresponding classical-quantum states. (See Theorem 1 of [35].)

1.2 Belavkin's Theorems

In the rest of this paper we shall restrict consideration to the ensemble

$$\mathcal{E}_m = \{(\psi_k, p_k)\}_{k=1, \dots, m} \quad (2)$$

of pure quantum states $\psi_k \in \mathcal{H}$, and consider POVMs given by

Definition 2 The *Belavkin Weighted Square Root Measurement (BWSRM)* [14, 36] (also known as a **Weighted Least-Squares Measurement** [37]) with weights $W_k \geq 0$ is the POVM¹

$$M_k = \left(\sum_{\ell} W_{\ell} |\psi_{\ell}\rangle \langle \psi_{\ell}| \right)^{-1/2} W_k |\psi_k\rangle \langle \psi_k| \left(\sum_{\ell} W_{\ell} |\psi_{\ell}\rangle \langle \psi_{\ell}| \right)^{-1/2}. \quad (3)$$

on the linear span of the $W_k |\psi_k\rangle$.

The importance of BWSRMs in minimum-error discrimination problems was shown by the following

¹The negative fractional power is well-defined on the restriction to the span of the $W_k \psi_k$. More properly, one may define $A^{-1/2} = \sum \lambda_k^{-1/2} |\phi_k\rangle \langle \phi_k|$, where $A = \sum \lambda_k |\phi_k\rangle \langle \phi_k|$ is a spectral-decomposition with $\lambda_k = 0$ terms omitted.

Theorem 3 (Belavkin 1975 [14, 36]) *A POVM $\{M_k\}$ on $\text{Span}(\mathcal{E}_m) \equiv \text{Span}(\{p_k \psi_k\})$ is optimal if and only if it may be expressed as a BWSRM with weights W_k such that the operator*

$$\Lambda = \left(\sum_{\ell=1}^m W_\ell |\psi_\ell\rangle \langle \psi_\ell| \right)^{1/2} \quad (4)$$

is invertible on $\text{Span}(\mathcal{E}_m)$ and

$$p_k \langle \psi_k | \Lambda^{-1} | \psi_k \rangle \leq 1, \quad (5)$$

with equality when $W_k > 0$.²

Note that a simple formula for a set of optimal weights corresponding to a given optimal measurement

$$W_k^{\text{opt}} = \langle \psi_k | M_k^{\text{opt}} | \psi_k \rangle \times p_k^2 \quad (6)$$

follows by squaring both sides of (5) and multiplying by W_k , whether or not $W_k = 0$. Furthermore, Belavkin's theorem implies that optimal measurements on $\text{Span}(\mathcal{E}_m)$ satisfy $\text{Rank}(M_k^{\text{opt}}) \leq 1$.³

Belavkin and Maslov generalized Theorem 3 to mixed states (and more generally to wave pattern recognition) in section 2.2 of [29]. Iteration of a mixed state version of equation (6) was explored in [27, 28] as a method for numerical computation of optimal measurements.

Many of the known exactly solvable optimal pure state measurements are special cases of

Theorem 4 (Belavkin 1975 [14, 36], also Ban [41]) *The measurement (3) for the weights $W_k = p_k$ is optimal for the pure state ensemble \mathcal{E}_m if the chance of successfully identifying a given state ψ_k is inversely proportional to its a priori probability.⁴*

$$p_k \langle \psi_k | M_k | \psi_k \rangle = \text{const.} \quad (7)$$

Note that condition (7) is sufficient but not necessary, as may be seen by considering direct sums. Belavkin originally applied Theorem 4 to homogeneous systems, cyclic systems, and systems of coherent states.[36] (A generalization of cyclic systems appears in [42]).

1.3 Sub-optimal measurements

In abstract studies of quantum channel capacities or quantum algorithms, numerical routines for solving specific instances of optimal measurement problem are often neither feasible nor desirable: one often has to rely on sub-optimal measurements. Several extant approximately optimal measurements are examples of

Definition 5 *For $r > 0$, the **Belavkin power-weighted square-root measurement (BWSRM- r)** is the BWSRM with weights $W_k = p_k^r$.*

Examples of BWSRM- r 's appearing in the literature correspond to $r = 1, 2, 3$. Note that in the case of equiprobable ($p_k = 1/m$) pure states that all BWSRM- r 's are identical, and are of pervasive utility in quantum information theory. (See, for example [2].)⁵

²Mochon rediscovered that every optimal pure-state measurement may be expressed as a BWSRM. [38]

³For mixed states, $\text{Rank}(M_k^{\text{opt}}) \leq \text{Rank}(\rho_k)$ [14]. See also equation 5 of [27] and [17]. Equation 6 may be understood geometrically using "frame forces," which have been advocated by Kebo and Benedetto [39, 40].

⁴The weighted measurement defined by $W_k = p_k$ sometimes appears as $M_k = |e_k\rangle \langle e_k|$, with $|e_k\rangle = \sum_{\ell=1}^m |\psi_\ell\rangle (P^{-1/2})_{\ell k}$, where P is the Graham matrix $P_{ij} = \sqrt{p_i p_j} \langle \psi_i, \psi_j \rangle$. Condition (7) is then written as $(\sqrt{P})_{ii} = (\sqrt{P})_{jj}$ for all i, j . The equivalence of these two formulations follows from the matrix identities $\Gamma^\dagger (\Gamma \Gamma^\dagger)^{-1/2} = (\Gamma^\dagger \Gamma)^{-1/2} \Gamma^\dagger$ and $\Gamma (\Gamma^\dagger \Gamma)^{-1/2} \Gamma^\dagger = (\Gamma \Gamma^\dagger)^{1/2} = \sqrt{P}$, where $\Gamma = \sum_{\ell=1}^m \sqrt{p_\ell} |\ell\rangle_{\mathbb{C}^m} \langle \psi_k |_{\mathcal{H}} : \mathcal{H} \rightarrow \mathbb{C}^m$. Here $\{|\ell\rangle_{\mathbb{C}^m}\}$ is the standard orthonormal basis of \mathbb{C}^m .

⁵The study of BWSRM- r 's as approximately-optimal measurements in the equiprobable case goes as far back as [1] and [43].

We have already encountered the $r = 1$ case in Theorem 4. This measurement came to be known as the “pretty good measurement,” (PGM) because of its reintroduction two decades later by Hausladen and Wootters as an *ad hoc approximately optimal measurement* [44, 45] *with simple error bounds*. Barnum and Knill showed that the failure rate of the mixed-state version

$$M_k^{\text{PGM}} = \left(\sum p_\ell \rho_\ell \right)^{-1/2} p_k \rho_k \left(\sum p_\ell \rho_\ell \right)^{-1/2}$$

of the Belavkin-Hausladen-Wootters PGM satisfies the bound

$$P_{\text{fail}}^{\text{opt}} \leq P_{\text{fail}}^{\text{PGM}} \leq P_{\text{fail}}^{\text{opt}} (1 + P_{\text{succ}}^{\text{opt}}) \leq 2P_{\text{fail}}^{\text{opt}}, \quad (8)$$

where $P_{\text{fail}}^{\text{opt}}$ is the minimum-error failure rate. [19, 20] The bound

$$P_{\text{fail}}^{\text{PGM}} \leq \sum_{i \neq j} p_i |\langle \psi_i, \psi_j \rangle|^2 \quad (9)$$

was proved by Hayden *et al* [18], generalizing the equiprobable bound of [2].⁶

The cube-weighted BWSRM-3 was employed by Ballester, Wehner, and Winter in the study of state discrimination with post-measurement information.[46, 47]

1.4 Asymptotically-optimal measurements & BWSRM-2

The quadratically weighted BWSRM-2 will be of particular interest in the present work, and its mixed state generalization will be studied in the sequel. Relatively recently, the mixed state version of BWSRM-2 has appeared as the first iteration in a sequence of closed form measurements which appear to converge to the optimal measurement.[27, 28]. This weighting was first specifically considered by Holevo, who was most interested in the case of nearly orthogonal ψ_k .

Definition 6 *A measurement procedure G for distinguishing the pure-state ensemble \mathcal{E}_m is **asymptotically optimal** [48] if for fixed p_1, \dots, p_m one has*

$$\frac{P_{\text{fail}}^G(\mathcal{E}_m)}{P_{\text{fail}}^{\text{opt}}(\mathcal{E}_m)} \rightarrow 1$$

*as the states ψ_k approach an orthonormal basis.*⁷

Holevo showed that

Theorem 7 (Holevo’s asymptotic-optimality Theorem (1977) [48]) *The quadratically-weighted pure state Belavkin measurement BWSRM-2 is asymptotically optimal.*

As we will see in section 2.2, this property is **not** shared by the “pretty good measurement.” The key idea in Holevo’s proof was the construction of BWSRM-2 using an approximate minimal principle:

Theorem 8 (Holevo 1977 [48].) *Assume that the states ψ_k are linearly independent. Then the von Neumann measurement $M_k = |e_k\rangle\langle e_k|$ minimizing*

$$C^{\text{Holevo}}(\{e_k\}) = \sum_{k=1}^m p_k \|\psi_k - e_k\|^2 \quad (10)$$

over orthonormal⁸ sets $\{e_k\}$ is the quadratically weighted Belavkin measurement BWSRM-2.

⁶Equation (9) follows by summing the conditional error bound (A6) of [18]. Bounds based on the pairwise quantities $|\langle \psi_i, \psi_j \rangle|^2$ are inherently limited [20], although frequently useful.

⁷It is presumably intractable to produce a closed-form measurement process G for which $P_{\text{fail}}^G(\mathcal{E}_m)/P_{\text{fail}}^{\text{opt}}(\mathcal{E}_m) \rightarrow 1$ as the ψ_k and p_k are arbitrarily varied in such a way that $P_{\text{fail}}^{\text{opt}}(\mathcal{E}_m) \rightarrow 0$. Otherwise, one could recover the optimal measurement for a fixed ensemble \mathcal{E}_m on \mathcal{H} by taking the $\lambda \rightarrow 1^-$ limit of the ensemble $\mathcal{E}'_{m+1} \equiv \{(\psi_k, (1-\lambda)p_k)\} \cup \{(\phi, \lambda)\}$ on a dilation $\mathcal{H}' \supset \mathcal{H}$, with $\phi \perp \mathcal{H}$.

⁸Orthogonal measurements are optimal for distinguishing linearly-independent pure states.[1, 36, 38, 49]

Theorem 8 was generalized by Eldar and Forney [37], who showed that the BWSRM with weights W_k minimizes $C^{\{W_k\}} = \sum W_k \|\psi_k - e_k\|^2$ over POVMs $M_k = |e_k\rangle\langle e_k|$ without any assumption of linear independence.⁹

1.5 Results

In section 2.3 it is shown that a weighted measurement is asymptotically optimal only if it is quadratically weighted, proving a converse to Holevo's asymptotic optimality theorem. In section 2.2 the PGM is found to be categorically worse than the quadratically weighted measurement for two pure states. In section 2.4 we make a heuristic comparison between various weightings, and present a counter-example to show that the relationship between weightings is more complicated for ensembles of more than two states. Finally, in section 2.5 we compare sufficient optimality conditions for various weightings.

2 Pure State weighted measurements

2.1 Continuity

Although weighted measurements are defined using the singular map $x \mapsto x^{-1/2}$, one still has

Theorem 9 *For fixed weights W_k , $k = 1, \dots, m$, the success rate of the weighted measurement for distinguishing the pure state ensemble $\mathcal{E}_m = \{(\psi_k, p_k)\}_{k=1, \dots, m}$ is a jointly continuous function of the ψ_k and p_k .*

Proof. Define the operator $A : \mathbb{C}^m \rightarrow \mathcal{H}$ by

$$A = \sum_{k=1}^m \sqrt{W_k} |\psi_k\rangle_{\mathcal{H}} \langle k|_{\mathbb{C}^m}, \quad (11)$$

where $\{|k\rangle_{\mathbb{C}^m}\}$ is the standard orthonormal basis of \mathbb{C}^m . Then

$$\begin{aligned} P_{\text{succ}}^{W\text{-weighted}} &= \sum_{k=1}^m p_k \left| \langle \psi_k | \left(\sum_{\ell=1}^m W_{\ell} |\psi_{\ell}\rangle \langle \psi_{\ell}| \right)^{-1/2} W_k^{1/2} |\psi_k\rangle \right|^2 \\ &= \sum_{k=1}^m \frac{p_k}{W_k} \left| \langle k|_{\mathbb{C}^m} A^{\dagger} (A A^{\dagger})^{-1/2} A |k\rangle_{\mathbb{C}^m} \right|^2 \\ &= \sum_{k=1}^m \frac{p_k}{W_k} \left(\langle k| (A^{\dagger} A)^{+1/2} |k\rangle \right)^2. \end{aligned} \quad (12)$$

Continuity of P_{fail} follows from the continuity of the square root.¹⁰ ■

⁹One can recover this generalization from Holevo's argument using Naimark's Theorem. [39] Note that the cost function C^{Holevo} for arbitrary p_k already appears as eq. 8 of [48].

¹⁰By the Weierstrauss approximation theorem [50], given $\varepsilon > 0$ one can find a polynomial P such that $|P(\lambda) - \sqrt{\lambda}| < \varepsilon$ for all λ in the interval $I = [0, \sum W_k]$. Since I contains the spectrum of $A^{\dagger} A$ for any choice of $\{\psi_k\}$, continuity of (12) is guaranteed by Theorem 7.12 of [50].

2.2 Explicit comparison of weighted Belavkin measurements for 2 pure states

We first consider binary ensembles:

Theorem 10 *The failure rates for distinguishing the binary ensemble \mathcal{E}_2 using optimal and weighted measurements are given by*

$$P_{fail}^{optimal} = \frac{1}{2} - \sqrt{\frac{1}{4} - p_1 p_2 |\langle \psi_1, \psi_2 \rangle|^2} \quad (13)$$

$$P_{fail}^{weighted} = \frac{(p_1 W_2 + p_2 W_1) \cos^2 \theta}{W_1 + W_2 + 2\sqrt{W_1 W_2} |\sin \theta|}, \quad (14)$$

where $\cos \theta = |\langle \psi_1, \psi_2 \rangle|$.

Proof. Equation (13) is equation 2.34 on page 113 of [1].

For an arbitrary 2×2 positive matrix B it is easy to use the spectral theorem to verify that

$$B^{+1/2} = \left(2\sqrt{\det B} + \text{Tr } B\right)^{-1/2} \left(B + \sqrt{\det B} \times \mathbb{1}\right).$$

For A defined by (11) one has

$$\det A^\dagger A = W_1 W_2 \sin^2 \theta$$

$$\text{Tr } A^\dagger A = W_1 + W_2.$$

Equation (14) now follows from (12):

$$\begin{aligned} P_{\text{succ}}^{W_1, W_2} &= \sum \frac{p_k}{W_k} \left| \langle k | (A^\dagger A)^{+1/2} | k \rangle \right|^2 \\ &= \sum \frac{p_k}{W_k} \left| \left(2\sqrt{W_1 W_2} |\sin \theta| + W_1 + W_2\right)^{-1/2} \left(W_k + \sqrt{W_1 W_2} |\sin \theta|\right) \right|^2 \\ &= 1 - \frac{\sum p_k W_{1-k} \cos^2 \theta}{W_1 + W_2 + 2\sqrt{W_1 W_2} |\sin \theta|}. \end{aligned}$$

■

Theorem 11 (Holevo's measurement is better than the PGM for two pure states) *For distinguishing the 2-pure-state ensemble \mathcal{E}_2 one has the following inequalities*

$$P_{fail}^{PGM} \geq P_{fail}^{Holevo}, \quad (15)$$

$$2 = \sup_{\mathcal{E}_2} \frac{P_{fail}^{PGM}}{P_{fail}^{optimal}} > \sup_{\mathcal{E}_2} \frac{P_{fail}^{Holevo}}{P_{fail}^{optimal}} = \frac{\sqrt{2} + 1}{2} \approx 1.207 \quad (16)$$

$$> \sup_{\mathcal{E}_2} \frac{P_{fail}^{cubic \text{ weighting}}}{P_{fail}^{optimal}} \approx 1.118, \quad (17)$$

with equality in (15) iff $p_1, p_2 \in \{0, 1/2, 1\}$ or $\langle \psi_1, \psi_2 \rangle = 0$.

Proof. To prove (15), note that since $\sqrt{p_1 p_2} \leq \frac{1}{2}$, we have the inequalities

$$\begin{aligned} \frac{1}{2} + \sqrt{p_1 p_2} &\leq 1 \\ \sqrt{p_1 p_2} (|\sin \theta| - 1) &\leq 2p_1 p_2 (|\sin \theta| - 1). \end{aligned}$$

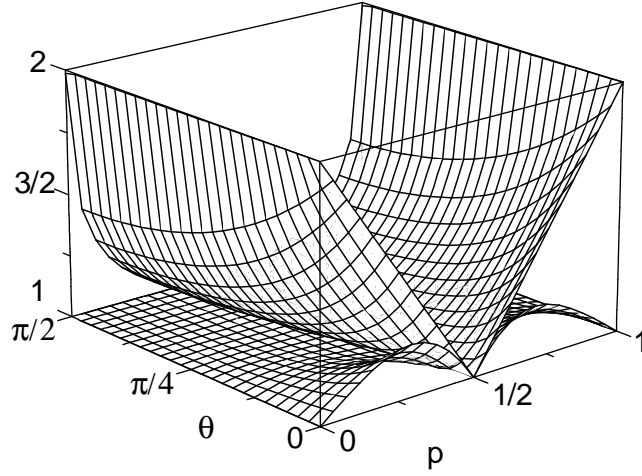


Figure 1: $P_{\text{fail}}/P_{\text{fail}}^{\text{opt}}$ for the PGM (upper) and Holevo's measurement (lower) for binary ensembles with $|\langle\psi_1, \psi_2\rangle| = \cos\theta$ and $p_1 = 1 - p_2 = p$

Summing gives

$$\frac{1}{2} (1 + 2\sqrt{p_1 p_2} |\sin\theta|) \leq 1 + 2p_1 p_2 (|\sin\theta| - 1) = p_1^2 + p_2^2 + 2p_1 p_2 \sin|\theta|.$$

Equation (15) follows by dividing $p_1 p_2 \cos^2\theta$ by both sides and applying (14).

The equation on the left-hand side of (16) shows that the bound (8) of Barnum and Knill is sharp. To see that

$$\sup_{\mathcal{E}_2} \frac{P_{\text{fail}}^{\text{PGM}}}{P_{\text{fail}}^{\text{optimal}}} \geq 2,$$

take $p_1 \rightarrow 0^+$ for any fixed $\langle\psi_1, \psi_2\rangle \neq 0$. The equation on the RHS of (16) is an unilluminating exercise in calculus. The maximizing ensemble is given by $\psi_1 = \psi_2$ and $p_1 = \sqrt{2}/2$. The last inequality (17) was computed numerically. ■

The relative success rates of $P_{\text{fail}}/P_{\text{fail}}^{\text{optimal}}$ for the weightings $r = 1, 2$, and 3 of measurements on the ensemble \mathcal{E}_2 with $|\langle\psi_1, \psi_2\rangle| = \cos\theta$ and $p_1 = 1 - p_2 = p$ are plotted in Figures 1 and 2a/b.

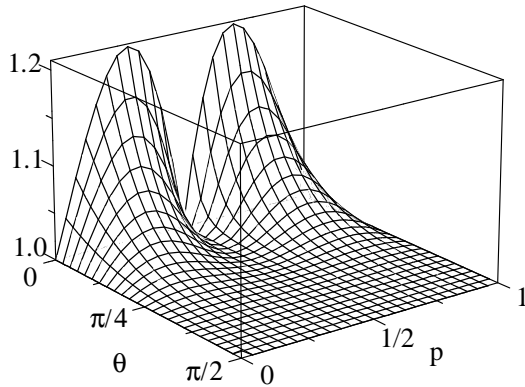


Fig 2(a): $P_{\text{fail}}^{\text{Holevo}}/P_{\text{fail}}^{\text{opt}}$ for binary ensembles

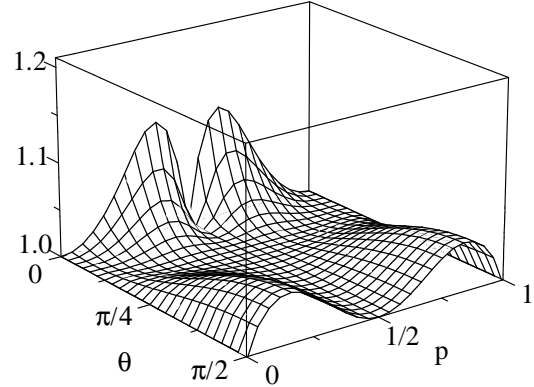


Fig 2(b): $P_{\text{fail}}^{\text{cubic}}/P_{\text{fail}}^{\text{opt}}$ for binary ensembles

2.3 Asymptotic optimality

Holevo's quadratic weighting is uniquely characterized by the following converse of Theorem 7:

Theorem 12 (Converse to Holevo's asymptotic optimality Theorem) *Fix probabilities $p_k > 0$ and weights $W_k \geq 0$, $k = 1, \dots, m$. Then the Belavkin weighted square root measurement (\mathfrak{J}) is asymptotically optimal for distinguishing ensembles $\mathcal{E}_m = \{(\psi_k, p_k)\}_{k=1, \dots, m}$ only if $W_k = \text{const} \times p_k^2$.*

Proof. Setting $c_k = W_k/p_k^2$, we must show that $c_k = c_{k'}$ for all k, k' under the assumption that $\{W_k\}$ defines an asymptotically optimal measurement. It is sufficient to consider the case $m = 2$.¹¹ By Theorem 10 and L'Hospital's rule

$$\begin{aligned} \lim_{\theta \rightarrow \pi/2} \frac{P_{\text{fail}}^{\text{weighted}}}{P_{\text{fail}}^{\text{optimal}}} &= \frac{p_1 W_2 + p_2 W_1}{W_1 + W_2 + 2\sqrt{W_1 W_2}} \times \lim_{\theta \rightarrow \pi/2} \frac{\cos^2 \theta}{\frac{1}{2} - \sqrt{\frac{1}{4} - p_1 p_2 \cos^2 \theta}} \\ &= \frac{(c_2 p_2 + c_1 p_1) p_1 p_2}{(\sqrt{W_1} + \sqrt{W_2})^2} \times \frac{1}{p_1 p_2} \\ &= \frac{c_1 p_1 + c_2 p_2}{(\sqrt{c_1 p_1} + \sqrt{c_2 p_2})^2}. \end{aligned}$$

The conclusion follows from the strict convexity of $x \mapsto x^2$. ■

2.4 Reflections & Counter-examples for three states

We now reflect on the relationships between Belavkin's optimal weighting ($W_k = p_k^2 \langle \psi_k | M_k^{\text{opt}} | \psi_k \rangle$), the weighting for the PGM ($W_k = p_k$), Holevo's weighting ($W_k = p_k^2$), and the weighting of Ballester, Wehner, and Winter ($W_k = p_k^3$). Note that while Holevo's measurement relatively over-weights vectors ψ_k for which $\langle \psi_k | M_k^{\text{opt}} | \psi_k \rangle$ is relatively small, the PGM additionally over-weights vectors for which p_k is small! In general, one therefore expects that *the relative misweightings of the PGM tend to compound one another, so that BWSRM-2 is better than BWSRM-1*. Similarly, by approximate cancellation of misweightings, one expects that the cubic weighting will sometimes outperform the quadratic weighting for ensembles far from the asymptotically orthogonal regime considered by Holevo.

We have seen that Holevo's measurement is always as least as good as the PGM for two-state ensembles. For three states the above intuitive argument does not always hold true, as shown by the following pathology:

Theorem 13 *There exists a 3-state ensemble with the properties that:*

1. *There is an optimal measurement such that the a priori strictly-most-probable state is NEVER detected.*
2. *Holevo's measurement is worse than the PGM: $P_{\text{fail}}^{\text{Holevo}} > P_{\text{fail}}^{\text{PGM}}$.*

Proof. Define the 3-state ensemble by $\psi_1 = (\cos \theta, \sin \theta)$, $\psi_2 = (\cos \theta, -\sin \theta)$, and $\psi_3 = (1, 0)$, where $\theta = \pi/6$ and

$$p_1 = p_2 = (1 - p_3)/2 = (2 + (\cos \theta + \sin \theta) \cos \theta)^{-1} \approx .3142 < 1/3.$$

It is straightforward to check that the POVM

$$M_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad M_2 = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad M_3 = 0$$

¹¹The case $m > 2$ is reduced to $m = 2$ by considering ensembles for which each of a subset $m-2$ states is orthogonal to all of the other states in \mathcal{E}_m .

satisfies the necessary and sufficient optimality conditions [13]

$$(L + L^\dagger) / 2 - p_k |\psi_k\rangle \langle \psi_k| \geq 0 \text{ for all } k,$$

where the Lagrange operator L is given by

$$L \equiv \sum_{k=1}^3 p_k M_k |\psi_k\rangle \langle \psi_k| = p_1 (\cos(\theta) + \sin(\theta)) \begin{bmatrix} \cos \theta & \\ & \sin \theta \end{bmatrix}.$$

(Here $A \geq 0$ means A is positive semidefinite.) Property 2 follows by direct computation:

$$P_{\text{fail}}^{\text{Holevo}} \approx .4245 > P_{\text{fail}}^{\text{PGM}} \approx .4224 > P_{\text{fail}}^{\text{optimal}} \approx .4138.$$

■

Remark: The linear-dependence of the states in the above construction was not essential: one can simply embed the above example in 3-space, and perturb the vectors ψ_k slightly to make them linearly independent. By Theorem 9, property 2 will be unaffected by small perturbations.

Theorem 13 aside, given any fixed set of non-equal priors p_k , we conjecture that Holevo's weighting will have a better success rate than the PGM on average for randomly chosen ensembles \mathcal{E}_m , with the corresponding $\{\psi_k\}$ independently chosen according to Haar measure.

2.5 Sufficient optimality conditions for weighted measurements

We close our comparisons of Belavkin weighted measurements by noting that in the case $W_k > 0$, Theorem 4 generalizes easily:

Theorem 14 (Optimality conditions for positively-weighted measurements) *A sufficient condition for optimality of the BWSRM with strictly positive weights $W_k > 0$ is that there exists a constant $c > 0$ such that*

$$p_k^2 \langle \psi_k | M_k | \psi_k \rangle = c W_k \text{ for all } k. \quad (18)$$

Proof. Dividing both sides of (18) by $c W_k$ and taking the square root gives

$$c^{-1/2} p_k \langle \psi_k | \left(\sum W_\ell |\psi_\ell\rangle \langle \psi_\ell| \right)^{-1/2} | \psi_k \rangle = 1.$$

In particular, the rescaled weights $c \times W_\ell$ satisfy Belavkin's optimality condition (5). The result follows, since BWSRMs are unaffected by such rescalings. ■

The assumption that $W_k > 0$ for all k is necessary: otherwise the weights $W_k = \delta_{k1}$ would be optimal for any ensemble. Note that for the asymptotically optimal weight $W_k = p_k^2$, equation (18) becomes the particularly simple condition

$$\langle \psi_k | M_k | \psi_k \rangle = \text{const}. \quad (19)$$

3 Future directions

In the sequel, we focus on the quadratically weighted mixed state measurement, and consider resulting two-sided bounds for the distinguishability arbitrary ensembles of mixed quantum states [51]. We will generalize to the case of approximate reversals of quantum channels at a later date [52].

Acknowledgements: I would like to thank Andrew Childs, Aram Harrow, Julio Concha, V. P. Belavkin, and Vincent Poor for pointing out useful references, Julio Concha and Andrew Kebo for providing copies of their theses, William Wootters for providing a copy of Hausladen's thesis, and Arthur Jaffe for his encouragement.

References

- [1] Carl W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, New York 1976.
- [2] P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, and W. K. Wootters, “Classical information capacity of a quantum channel,” *Physical Review A* **54**, pp. 1869-1876 (1996).
- [3] Benjamin Schumacher and Michael D. Westmoreland, “Sending classical information via noisy quantum channels,” *Physical Review A* **56** 131-138 (1997).
- [4] A. S. Holevo, “The capacity of the quantum channel with general signal states,” *IEEE Transactions on Information Theory* **44** pp. 269-273 (1998).
- [5] Lawrence Ip. “Shor’s algorithm is optimal,” Preprint, 2004; <http://lawrenceip.com/papers/hspdpabstract.html>.
- [6] D. Bacon, A. M. Childs, and W. van Dam, “Optimal measurements for the dihedral hidden subgroup problem,” *Chicago Journal of Theoretical Computer Science* (2006), no.2; quant-ph/0501044.
- [7] Dave Bacon, Andrew M. Childs, and Wim van Dam, “From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups,” *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)* (Society of Industrial and Applied Mathematics, Philadelphia 2007) pp. 469-478; quant-ph/0504083.
- [8] Andrew Childs and Wim van Dam, “Quantum algorithm for a generalized hidden shift problem,” *Proc. 18th ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*, pp. 1225-1234; quant-ph/0507190.
- [9] Cristopher Moore and Alexander Russell, “For Distinguishing Hidden Subgroups, the Pretty Good Measurement is as Good as it Gets,” *Quantum Information and Computation* **7** 752-765 (2007); quant-ph/0501177.
- [10] Dave Bacon and Thomas Decker, “The optimal single-copy measurement for the hidden-subgroup problem,” *Physical Review A* **77**, 032335 (2008); arxiv:0706.4478.
- [11] Horace P. Yuen, Robert S. Kennedy, and Melvin Lax, “Optimum testing of multiple hypotheses in quantum detection theory,” *IEEE Transactions on Information Theory*, **IT-21**, no. 2, 125-134 (March 1975).
- [12] A. S. Holevo, “Statistical Decision Theory for Quantum Systems,” *Journal of Multivariate Analysis* **3**, pp. 337-394 (1973).
- [13] A. S. Holevo, “Remarks on optimal measurements,” *Problemy Peredachi Informatsii*, **10** no. 4, 51-55; English Translation: *Problems of Information Transmission* **10**, no.4, 317-320 (1974).
- [14] V. P. Belavkin, “Optimal multiple quantum statistical hypothesis testing.” *Stochastics* **1**, 315-345 (1975). Note: Inequality 4.3 in the statement of Theorem 5 is backwards.
- [15] V. P. Belavkin and A. G. Vancjan, *Radio Engineering and Electronic Physics* **19** (7), 1397-1401 (1974).
- [16] S. M. Barnett and S. Croke, “On the conditions for discrimination between quantum states with minimum error,” *J. Phys. A: Math. Theor.* **42** 062001 (2009); arxiv:0810.1919.
- [17] Yonina C. Eldar, Alexandre Megretski, and George C. Verghese, “Designing Optimal Quantum Detectors Via Semidefinite Programming,” *IEEE Transactions on Information Theory*, Vol 49 #4, pp. 1007-1012 (2003).

- [18] Patrick Hayden, Debbie Leung, and Graeme Smith, “Multiparty data hiding of quantum information,” *Physical Review A* **71**, 062339 (2005); quant-ph/0407152.
- [19] H. Barnum and E. Knill, “Reversing quantum dynamics with near-optimal and classical fidelity,” *Journal of Mathematical Physics* **43** (5), pp. 2097-2016 (2002); quant-ph/0004088.
- [20] Ashely Montanaro, “On the distinguishability of random quantum states,” *Communications in Mathematical Physics* **273**, 619-636 (2007); quant-ph/0607011.
- [21] Daowen Qiu, “Minimum-error discrimination between mixed quantum states,” *Physical Review A* **77**, 012328 (2008); arXiv:0707.3970.
- [22] Ashley Montanaro, “A lower bound on the probability of error in quantum state discrimination,” Proc. IEEE Information Theory Workshop 2008. ITW ’08 (IEEE, Piscataway, NJ 2008), pp. 378-380; arXiv: 0711.2012.
- [23] Dawoen Qiu and Lvjun Li, “Bounds on the minimum-error discrimination between mixed quantum states,” arXiv:0812.2378.
- [24] Masahito Hayashi, Akinori Kawachi, and Hirotada Kobayashi, “Quantum measurements for hidden subgroup problems with optimal sample complexity,” *Quantum Information and Computation* **8**, 0345-0358 (2008); quant-ph/0604174.
- [25] J. Tyson, “Estimates of non-optimality of quantum measurements and a simple iterative method for computing optimal measurements,” arXiv:0902.0395.
- [26] Carl W. Helstrom, “Bayes-Cost reduction algorithm in quantum hypothesis testing,” *IEEE Transactions on Information Theory* **IT-28** no.2, March 1982 359-366.
- [27] M. Ježek, J. Řeháček, and J. Fiurášek, “Finding optimal strategies for minimum-error quantum state discrimination,” *Physical Review A* **65**, 060301(R) (2002); quant-ph/0201109.
- [28] Zdeněk Hradil, Jaroslav Řeháček, Jaromír Fiurášek, and Miroslav Ježek, “Maximum-Likelihood Methods in Quantum Mechanics,” *Lecture Notes in Physics* **649**, 59-112 (2004).
- [29] V. P. Belavkin and V. Maslov, “Design of Optimal Dynamic Analyzer: Mathematical Aspects of Wave Pattern Recognition” In *Mathematical Aspects of Computer Engineering*, Ed V. Maslov, pp. 146-237 (Mir, Moscow 1987); arXiv:quant-ph/0412031. **Note:** The first two equations on page 40 should be $F_i = H_i (L^\circ)^{-1/2}$ and $M_i^\circ = (L^\circ)^{+1/2} D_i^\circ (L^\circ)^{+1/2}$.
- [30] Andrew S. Fletcher, “Channel-Adapted Quantum Error Correction,” *PhD Thesis* MIT 2007; arxiv:0706.3400.
- [31] Andrew S. Fletcher, Peter W. Shor, and Moe Z. Win, “Optimum quantum error recovery using semidefinite programming,” *Physical Review A* **75**, 012338 (2007); quant-ph/0606035.
- [32] A. S. Fletcher, P.W. Shor, and M. Z. Win, “Channel-Adapted Quantum Error Correction for the Amplitude Damping Channel,” *IEEE Trans. Inf. Theory*, **54** 5705-5718 (2008); arXiv:0710.1052.
- [33] Andrew S. Fletcher, Peter W. Shor, and Moe Z. Win, “Structured near-optimal channel-adapted quantum error correction,” *Physical Review A* **77**, 012320 (2008); arxiv:0708.3658.
- [34] Soraya Taghavi, Robert L. Kosut, and Daniel A. Lidar, “Channel-Optimized Quantum Error Correction,” arXiv:0810.2524.
- [35] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” arXiv:0807.1338.

- [36] V. P. Belavkin, *Radiotekhnika i Elektronika* **20** 6, pp. 1177-1185 (1975). English translation: “Optimal distinction of non-orthogonal quantum signals,” *Radio Engineering and Electronic Physics* **20**, pp. 39-47 (1975).
- [37] Yonina C. Eldar and G. David Forney, “On quantum detection and the square-root measurement,” *IEEE Transactions on Information Theory* **47** 858-872 (2001); quant-ph/0005132.
- [38] Carlos Mochon, “Family of generalized ‘pretty good’ measurements and the minimal-error pure-state discrimination problems for which they are optimal,” *Physical Review A* **73**, 032328 (2006); quant-ph/0506061. (Note added to arXiv version: This paper has been withdrawn from the arXiv, because of similarity with the results of [36].)
- [39] Andrew Kei Kebo, “Quantum detection and finite frames,” *Ph.D dissertation* University of Maryland, College Park, 2005.
- [40] John J. Benedetto and Andrew Kebo, “The role of frame force in quantum detection,” *J. Fourier Anal Appl* (2008) **14**:443-474.
- [41] Masashi Ban, “Optimal signal detection in entanglement-assisted quantum communication systems,” *Journal of Optics B: Quantum and Semiclassical Optics* **4**, 143-148 (2002).
- [42] Tsuyoshi Sasaki Usuda, Ichi Takumi, Masayasu Hata, and Osamu Hirota, “Minimum error detection of a classical linear code sending through a quantum channel,” *Physics Letters A* **256** 104-108 (1999).
- [43] Paul Joseph Curlander, “Quantum Limitations on Communication Systems” *Ph.D. Thesis* MIT 1979.
- [44] P. Hausladen, “On the Quantum Mechanical Channel Capacity as a Function of the Density Matrix,” *B. A. Thesis*, Williams College, Williamstown, Massachusetts 1993.
- [45] P. Hausladen and W. K. Wothers, “A ‘pretty good’ measurement for distinguishing quantum states,” *Journal of Modern Optics* **41**, pp. 2385-2390 (1994).
- [46] Stephanie Wehner, “Cryptography in a quantum world,” *PhD Thesis: University of Amsterdam* (2008); arxiv:0806.3483.
- [47] Manuel A. Ballester, Stephanie Wehner, and Andreas Winter, “State discrimination with post-measurement information,” *IEEE Transactions on Information Theory* **54** (9). pp. 4183-4198; quant-ph/0608014.
- [48] A. S. Kholevo, *Teoriia Veroiatnostei I Ee Primeneniia* **23**(2) 429-432 (1978). English translation: “On asymptotically optimal hypothesis testing in quantum statistics,” *Theory of Probability and its Applications* **23**(2) 411-415 (1978). **Note:** The displayed equation between (8) and (9) should be $\sum_j \pi_j \|\psi_j - e_j\|^2 = 2(1 - \text{Re Tr}(U\Pi\Pi^{1/2}))$. The line just after equation (9) should read “where $V^* = |\Pi\Pi^{1/2}|(\Pi\Pi^{1/2})^{-1} \dots$ ”. The final expression in the paper should be $2(1 - \text{Tr}|\Pi^{1/2}\Pi|)$.
- [49] R. S. Kennedy, “On the optimum quantum receiver for the M-ary linearly independent pure state problem,” *MIT Research Laboratory of Electronics Quarterly Progress Report*, Technical Report No. 110, pp. 142-146, 1973.
- [50] Walter Rudin, *Principles of Mathematical Analysis*, 3rd Edition, McGraw-Hill 1976.
- [51] Jon Tyson, “Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo-Curlander bounds,” *J. Math. Phys.* **50**, 032106 (2009).
- [52] Jon Tyson, *in preparation*.